Supporting Information

Englert et al. 10.1073/pnas.1213795109

SI Methods

Plasmid Constructs and Protein Purification. For the first RNA-splicing ligase RtcB construct, 11-extra amino acid residues are AAALEHHHHHH. For the second N-Hig-Tev-RtcB construct, the tobacco etch virus (TEV) protease cleavage site is between Q and G in the MN-HHHHHH-ENLYFQ-GH-M sequence. After TEV protease cleaves, the RtcB protein has two extra amino acid residues GH before its first M.

For the first RtcB complex after lysis, the *Pyrococcus horikoshii* (*PH*) RtcB-1 supernatant in Ni-NTA buffer [20 mM Tris-HCl pH 8.0, 500 mM NaCl, 3 mM $MnCl_2$, 10 mM imidazole, 10 mM β-mercaptoethanol (β-ME)] was loaded on Ni-NTA column, washed with 40 mM imidazole, and eluted with 200 mM imidazole. The Ni-NTA elution fractions were directly loaded on a HiLoad Superdex 200 (26/60) gel-filtration column pre-equilibrated with 10 mM Tris-HCl pH 8.0, 200 mM NaCl, 3 mM $MnCl_2$, and 10 mM β-ME. The peak fractions with soluble monomeric *PH* RtcB-1 protein were concentrated to 10 mg/mL using an Amicon Ultra centrifugal device (molecular weight cut-off of 30,000) and frozen in liquid N_2 for storage. For enzymatic assays, the purified RtcB proteins were dialyzed against 20 mM Hepes-NaOH pH 7.9, 500 mM NaCl, 3 mM $MnCl_2$, and 10 mM β-ME. All buffer's pHs were adjusted and measured at 4 °C.

For the second RtcB complex after lysis, the *PH* RtcB-2 supernatant in a different Ni-NTA buffer (50 mM Na-phosphate pH 8.0, 500 mM NaCl, 60 mM KCl, 5% sucrose, and 5 mM β-ME) was loaded to 5 mL HisTrap HP (GE Healthcare) and eluted with a 25–400 mM imidizole gradient. The eluted RtcB fractions were treated with TEV protease (0.5 unit in 1-mg sample) to cleave the His-tag through overnight dialysis in buffer containing 20 mM Na-phosphate pH 7.5, 150 mM NaCl, 5% sucrose, 0.25 mM DTT, 1 mM EDTA, and 50 mM imidazole. The cleaved RtcB samples were centrifuged at 45,000 × *g* for 30 min to remove any precipitates. The cleared RtcB protein was further purified using a HiTrap Heparin HP (GE Healthcare) eluted with a 150–2,000 mM NaCl gradient in buffer containing 20 mM Hepes pH 7.5, 5% glycerol, and 1 mM DTT and using a Superdex 200 PG (26/60) gel-filtration column (Amersham) pre-equilibrated with 20 mM Hepes pH 7.5, 200 mM NaCl, 10% glycerol, and 1 mM DTT.

Ambiguity in Identification of Pyrophosphate or Trace Amount of Unreacted GTP in the RtcB-GMP/Mn^{2+} Complex. Next to the phosphate of 5′-GMP-His (GPH), there are two peaks in the residual Fo-Fc difference Fourier maps, which are too large and too far away from the enzyme, to be two ordered water molecules (Fig. 3 and Fig. S2). The spacing of these two peaks is consistent with the distance between two phosphate groups of pyrophosphate or the triphosphate moiety of GTP. When a pyrophosphate molecule was included in the refinement, a model suggests that it can form about 10 hydrogen bonds in the pocket near where SO_4-4 was observed in the RtcB/Mn^{2+} complex (Fig. 3). However, based on the distances of these peaks to the phosphate of GPH, GTP clearly fits better than pyrophosphate plus GPH (Fig. S2). It is possible that a minor fraction of unreacted GTP might bind at this site where H404 has to adopt a different rotameric conformation to avoid a steric clash.

A typical phosphoryl-transfer reaction involves concurrent bond breakage and bond formation with inversion of the phosphoryl center where the bonds to be broken and to be formed are linearly aligned. Although a minor fraction of unreacted GTP (or the pyrophosphate) may bind near the catalytic site, the orientation of the Pα-O3α bond, which is to be broken, is not aligned with the Pα-Nε2 bond of GPH (to H404), which has been formed as the product of the guanylylation reaction. Thus, we are unable to unambiguously establish the guanylylation reaction mechanism because the binding mode of GTP may not be catalytically relevant. The orientation of Pα-O3α of GTP and the binding of pyrophosphate predicts that H329 should be in a more optimal geometry for the enzyme guanylylation than H404. However, our mutational studies have ruled out any involvement of H329 in the enzyme guanaylylation step (Table S1), suggesting that the observed triphosphate moiety or the pyrophohsphate binding pocket is not catalytically relevant for the protein guanylylation; instead it is more relevant to the RNA guanylylation. In the H404 guanylylation reaction, the leaving pyrophosphate has to be located on the opposite side of the newly formed Pα-Nε2 bond of GPH. According to computer modeling, this pyrophosphate should bind to both Mn^{2+} ions observed in the RtcB/Mn^{2+} complex with the terminal γ-phosphate overlapping with SO_4-2 (Fig. S2), or bind near the SO_4-4 site next to R408/R412, which has yet to be experimentally resolved.

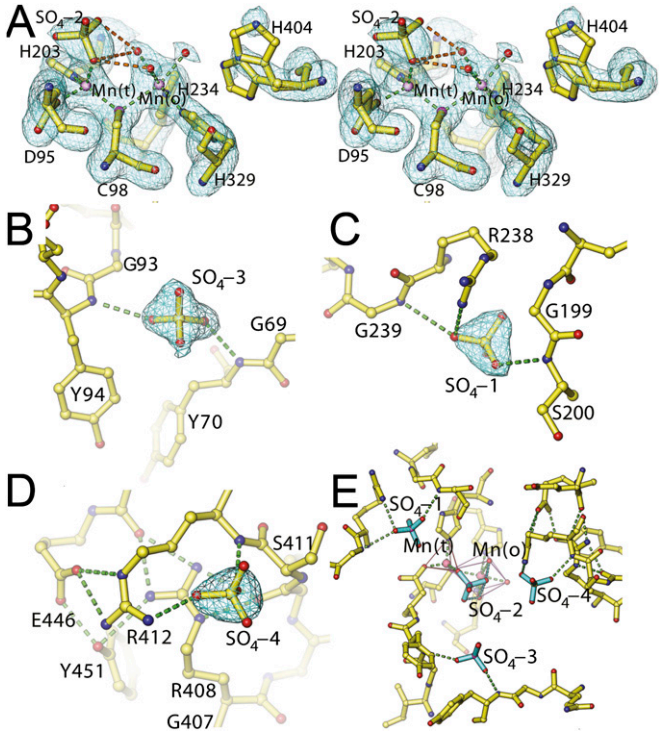

Fig. S1. The Mn^{2+}-bound RtcB complex. (*A*) Close-up view on the active site of RtcB showing two Mn^{2+} ions in tetraheadral and octahedral coordination by the indicated residues. A Fo-Fc difference Fourier residual map contoured at 2.5 σ superimposed on the refined model in stereo. (*B–D*) The binding residues for SO_4-3, SO_4-1, and SO_4-4 at corresponding sites superimposed onto Fo-Fc difference Fourier residual maps contoured at 3.5 σ, respectively. (*E*) Relationship among the four SO_4 binding sites to the two metal ions.

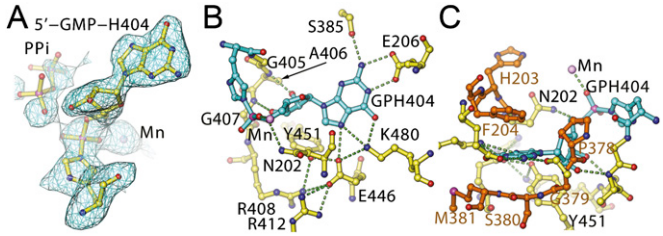

Fig. S2. Structure of RtcB-GMP/Mn^{2+} complex. (*A*) Fitting of GMP/PPi into a Fo-Fc difference Fourier residual map contoured at 3 σ. (*B*) Hydrogen bond interaction for GPH with indicated RtcB residues. (*C*) Hydrogen bonds (yellow) plus hydrophobic interactions (golden) for GPH.

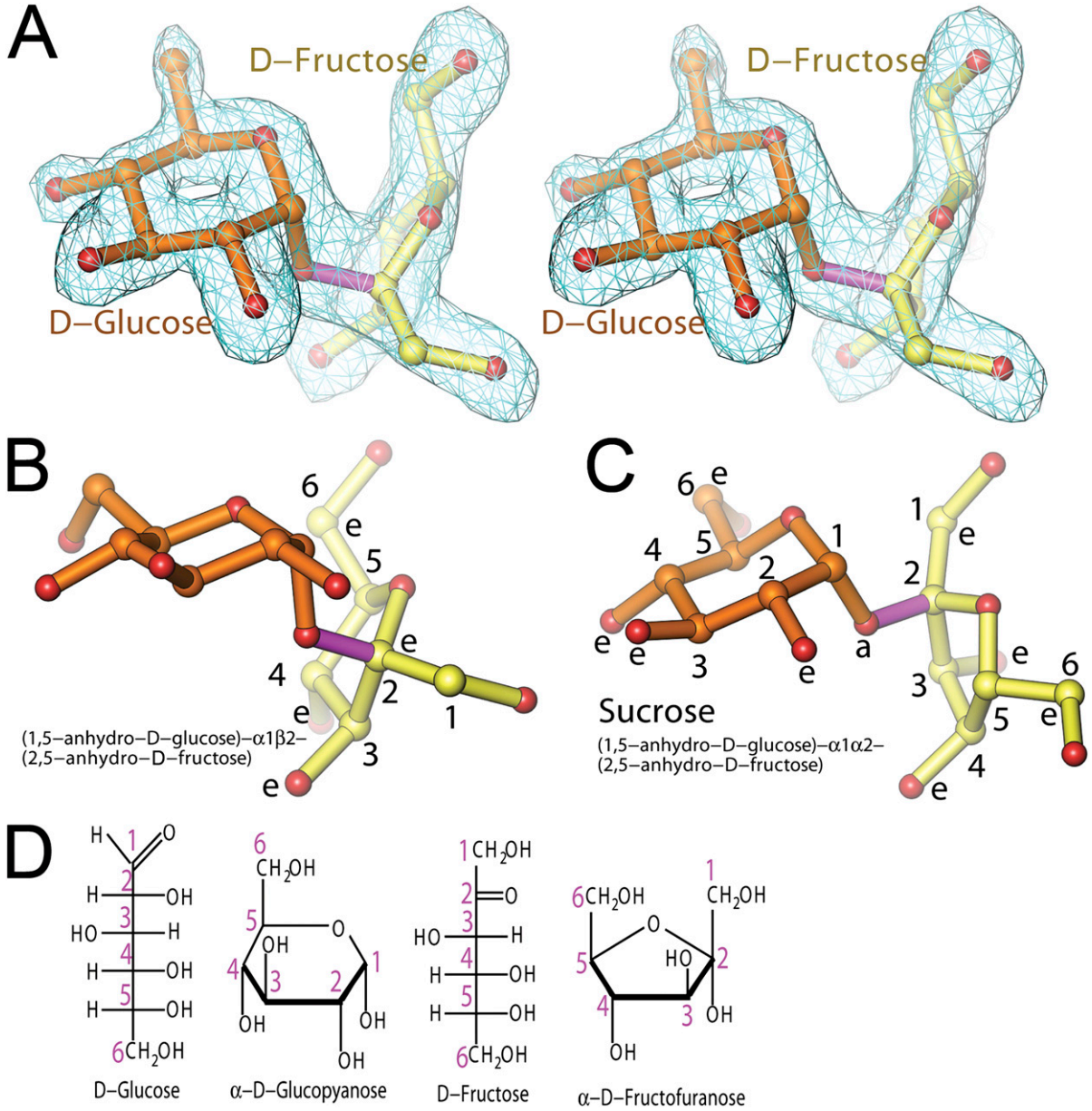

Fig. S3. Identification of unique sucrose. (*A*) An Fo-Fc difference Fourier residual map contoured at 3.5 σ superimposed onto the refined D-glucose-D-fructose disaccharide model in stereo. (*B*) Fitted D-glucose-D-fructose model in the structure. (*C*) Comparison of commonly occurring disaccharide, sucrose. (*D*) Fischer projection of D,D-glucose and fructose.

Table S1. Enzymatic characterization of conserved residues in RtcB

Mutations	RtcB guanyl.	RNA guanyl.	RNA ligation	Comments*
C98A	UD	UD	UD	Mn(o) and Mn(t) ligand
H329A	WD	7↓	50↓	Mn(o) ligand
H234A	1.5↑	2↑	7↓	Mn(o) ligand
D95	WT	UD	UD	Mn(t) ligand (SO_4-2 site)
H203A	3↓	1.5↑	UD	Mn(t) ligand (SO_4-2 site)
H404A	UD	UD	UD	The site of guanylylation
H404K	WD	UD	UD	The site of potential alternative guanylylation
D65A	UD	UD	UD	D65-H404 catalytic dyad for guanylylation
E206A	WD	UD	3↓	Recognition for G's N1 and N6
S385A	WD	UD	6↓	Recognition for G's N6
K480A	WD	UD	25↓	Recognition for G's O2
E446A	30↓	UD	6↓	Recognition for G's N7
F204A	WD	UD	50↓	Binding of G's nucleobase
Y451A	50↓	UD	UD	Binding of G's ribosyl and nucleobase moieties
R412A	10↓	UD	UD	SO_4-4 site; GTP's β,γ-phosphate sites
R408A	UD	UD	UD	Structural role; GTP's β,γ-phosphate sites
H67A	WT	1.5↑	7↓	The pyrophosphate site
N202A	WT	UD	UD	The GPH404's phosphate site
G199A	4↓	4↓	UD	5′-RNA with cyclic phosphate; SO_4-1 site
G239A	WT	WT	WT	5′-RNA with cyclic phosphate; SO_4-1 site
R238A	5↓	UD	UD	5′-RNA with cyclic phosphate; SO_4-1 site
R238A/G239A	2↓	UD	UD	5′-RNA with cyclic phosphate; SO_4-1 site
G69A	WT	WT	4↓	3′-RNA with free 5′-OH; SO_4-3 site
Y70A	WT	2↑	8↓	3′-RNA with free 5′-OH; SO_4-3 site
G69A/Y70A	3↓	WT	50↓	3′-RNA with free 5′-OH; SO_4-3 site
G93A	WT	WT	WT	3′-RNA with free 5′-OH; SO_4-3 site
Y94A	2↓	WT	WT	3′-RNA with free 5′-OH; SO_4-3 site
G93A/Y94A	WT	WT	4↓	3′-RNA with free 5′-OH; SO_4-3 site
F204Y	WD	3↓	6↓	Binding of G's nucleobase
P378G	2↓	WT	WT	Binding of G's nucleobase
P378G/F204Y	3↓	WT	2↓	Binding of G's nucleobase
N330A	2↓	1.5↑	WT	The pyrophosphate site
K351A	20↓	3.5↑	2↓	The pyrophosphate site

Assays for three enzymatic activities, RtcB guanylylation, 5′-RNA-3′-phosphate-guanylylation (5′-RNA-pp-G), and RNA ligation. UD, undetectable activity; WD, weakly detectable activity; WT, wild-type activity within experimental errors; ↓, decreased fold relative to the WT enzyme; ↑, increased fold relative to the WT enzyme.
*Structural features and plausible mechanistic implications.

Table S2. Summary of data processing and refinement statistics

RtcB complexes	RtcB/Mn^{2+} complex	RtcB-GMP/Mn^{2+} complex
X-ray data processing		
Space group	$P2_1$	$P2_12_12_1$
Unit cell dimensions (Å)	*a* = 114.08, *b* = 84.90, *c* = 124.20, β = 115.65°	*a* = 80.3, *b* = 137.8, *c* = 148.4
Wavelength (Å)	0.97918	0.978
Radiation Sources	24IDE, NECAT, APS	BL-1A, Photon Factory
Resolution (Å)*	50–1.55 (1.61–1.55)*	50–2.25 (2.33–2.25)*
No. of reflections	291,515 (19,787)	73,660 (7,320)
Redundancy	3.0 (1.9)	3.3 (3.3)
Completeness (%)	93.3 (63.6)†	93.8 (94.9)
$/\sigma_I$	18.7 (1.0)	10.5 (2.0)
R_{merge} (%)†	5.4 (61.6)	10.5 (48.9)
Refinement statisitcs		
ASU contents		
No. of residues	1,524	971
No. of water molecules	1,654	487
No. of Mn^{2+}	6	2
No. of atoms	13,603	8,195
No. of molecules	3	2
No. of reflections	274,606	69,979
Resolutions (Å)†,‡	47–1.55 (1.55–1.59)	101–2.25 (2.25–2.31)
$R_{working}$ (%)§	14.1 (38.3)	15.0 (23.1)
R_{free} (%)§	17.5 (42.9)	19.4 (30.4)
R_{o2p}¶	5.0	2.1
Rmsd‖		
Bonds (Å)	0.010	0.010
Bond angles (°)	1.32	1.63
PDB ID code	4DWR	4DWQ

*Statistics for the highest resolution (and the second highest) shell are in parenthesis. Data were proposed to the corners of detectors, where completeness decreased; and then reprocessed to the edge of detects to get the completeness information.

†$R_{merge} = \sum_{hkl}\sum_j |I_j(hkl) - \langle I_j(hkl)\rangle| / \sum_{hkl}\langle I(hkl)\rangle$, statistics for merging all observations for given reflections.

‡Although the effective resolution of the Mn^{2+}-bound structure is 1.55 Å, there was no reason to exclude data up to 1.49 Å resolution at the corner of the detectors (1). When all data at 1.49 Å resolution were included in refinement, the number of independent reflections increased to 294,058 and R_{o2p} increased to 5.4 with corresponding R factor and free R-factor of 14.2% and 17.8%, respectively.

§$R = \sum_{hkl}|F_{obs}(hkl) - F_{calc}(hkl)| / \sum_{hkl} F_{obs}(hkl)$, statistics for crystallographic agreement between the measured and model-calculated amplitudes. R_{free} is the agreement for the cross-validation dataset.

¶The observation-to-parameter ratio (R_{o2p}) is defined by the number of unique observation by the number of refinement parameters, which is four-times the number of atoms.

‖Root mean squares deviations (rmsd) to ideal values.

1. Wang J (2010) Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Crystallogr D Biol Crystallogr* 66:988–1000.